Designing an Empirical Framework to Estimate the Driver's Attention

Priyam Chowdhury, Lamia Alam, Mohammed Moshiul Hoque
Dept. of Computer Science & Engineering, Chittagong University of Engineering & Technology
Chittagong, Bangladesh
e-mail: priyamchowdhury70@gmail.com, lamiacse09@gmail.com, mmoshiulh@gmail.com

***Abstract*—Driver inattention is thought to cause many automobile crashes. Therefore, it is really important to pay high attention all the time though, for one split second of distraction can result in a terrible accident. However, developing an automatic driver assistance system for monitoring his/her attention during driving is a challenging task in computer vision. Level of attention of the driver may vary from high to low in different situations. Serious accident may be happened while the driver's attention becomes very low in certain cases. In this paper, we propose a vision based framework to estimate the level of attention of driver during driving. We have used Kinect sensor to collect the face angle and lip motions information to estimate level of attention. Moreover, the system generates an alarm as awareness signal of the driver in case of low attention level. On-board experiments demonstrate the effectiveness of the approach, even in case of adverse light conditions. Experiments with 10 driver's show that the system is functioning quite well and measure the driver's attention level with reasonable accuracy.**

Keywords—human computer interaction; computer vision; attention; level of attenton; head pose

I. INTRODUCTION

Most of the car crashes involve some form of distracted or inattentive driving. It is really important for drivers to pay attention all the time though, for one split second of distraction can result in an accident that causes destruction of lives and wealth. Level of attention may be low during driving for some reasons. For example, while drivers are involving in texting or talking over the cell phone, their eyes off the road and due to mind wandering. In case of a driver, it is an important issue to keep his/her attention level high while s/he is driving for the sake of his/her as well as the passenger's life. Meanwhile the level of attention of a driver is the measure of concentration while driving in terms of his/her physical and physiological parameters. In order to perform any kinds of work perfectly, humans need to perform attentively. But due to some unexpected situations and different physiological behavior, level of attention of a human is gradually goes down and as a result s/he cannot perform the task as s/he wanted to do. Sometimes lack of attention may be the threat for the human life. In case of driver, if s/he loses her/his attention while driving due to some physiological behaviors, a serious accident may happen within a second, which may cause serious injuries to the driver.

Driving is a complex task, requiring the concurrent execution of various cognitive, physical, sensory and psychomotor skills. Despite these complexities, it is not unusual to observe drivers engaging in various non-driving related activities while driving. These activities range from looking to other directions, conversing with passengers facing them. With the advent of wireless communication (e.g. mobile phones), more sophisticated entertainment systems and the introduction of technologies such as route navigation and the internet into vehicles, preoccupation with electronic devices while driving is also becoming increasingly common. Any activity that competes for the driver's attention while driving has the potential to degrade driving performance and have serious consequences for road safety. However, as more wireless communication, entertainment and driver assistance systems proliferate the vehicle market, it is likely that the rate of distraction related crashes will escalate. In particular, the focus is on the adaptive strategies drivers adopt when using devices in order to maintain their driving performance at an adequate level, under what conditions these adaptive strategies can fail and how driving performance is affected when they do. So driving with adequate level of attention plays a significant role in reducing the accident rate as well as assures safe journey. At the same time monitoring the driver's attention level is also an important issue for the improvement of the driving performance. This paper proposes a vision based system to monitor the level of attention of drivers in real time and generates a warning alarm while the level of attention is low as a safety measure.

II. RELATED WORK

There are lots of research activities have been conducted on designing an intelligent system related to safe driving. Although there are plenty of issues remained unsolved related to the driving. In this work, we would like to deal with an important capability of a system such as Driver's attention level estimation that is related to driving safely. Javier et al. [1] proposes a system uses a depth and visual information detect when a user is requiring attention and predict the objects of the environment that are suitable for interaction. The objective of this system is to detect the objects of the environment suitable for being used to interact with a human agent. Sun et al. [2] presented a hierarchical object-based visual attention system. Their strategies from course to fine occur on the multiple architecture of visual resolution and

groupings including objects, features and locations related to the relevant resolution. The model only considers covert attention (where the object does not move). In [3], to attract the attention and establish a communication channel based on VFOA proposed such a robotic system by developing computer vision methods to detect a target person's VFOA and its level. But here the attention level is measured based on face direction of the target parson only. Liyuan et al. [4] proposed a vision system to estimate attention of people from visual clues for social robot in interacting with multiple participants in public environments. Ba et al. [5] addressed the problem of recognizing the visual focus of attention of meeting participants based on their head pose. In order to recognize the human action using feature vectors that is derived from the skeleton model provided by the Kinect SDK in real time [6, 7]. Most of the previous works used standard color camera gave good results only in controlled environments. However, proposed system will work in real environment.

III. PROPOSED FRAMEWORK

Our main concern is to design a framework to measure the level of attention of the driver. Although there are lots of factors are involved in measuring the level of attention of drivers, in this work we consider the face angle and lip motion parameters. The driver should sit within the viable range of the Kinect sensor in order to collect the level of attention of him/her. RGB color camera of Kinect captures the driver. IR emitter emits dotted light patterns through the human and the pattern is read by the Kinect depth sensor. So a depth image of the driver is also produced. We used depth image for further processing. Based on the depth image initial position of the driver's head is determined. The later positioning is updated frame by frame according to the depth data which results in tracking the driver. The sensor sends continuous depth information to the Kinect device for further processing. Primary processing is done within the Kinect using the predefined algorithm. Then the depth information is sent to the personal computer for later processing. Then the face of the driver is tracked with the help of the development toolkit associated with the Kinect SDK [8, 9]. Using the animation units of the tracked face the lip motion can also be detected [10]. Level of attention is estimated by the predefined rules. The schematic representation of the proposed framework is illustrated in Fig. 1.

A. Capturing depth data

Kinect is consists of IR emitter, IR depth sensor and the color camera. The color camera is responsible for capturing and streaming the color video data and to detect there is a particle (i.e. human). The human is also captured by the depth camera and produce a depth image of that human. IR emitter and the IR depth sensor combined function to obtain the X, Y and Z coordinate values on specific point of human. We take the head part of the human body into consideration in the current implementation.

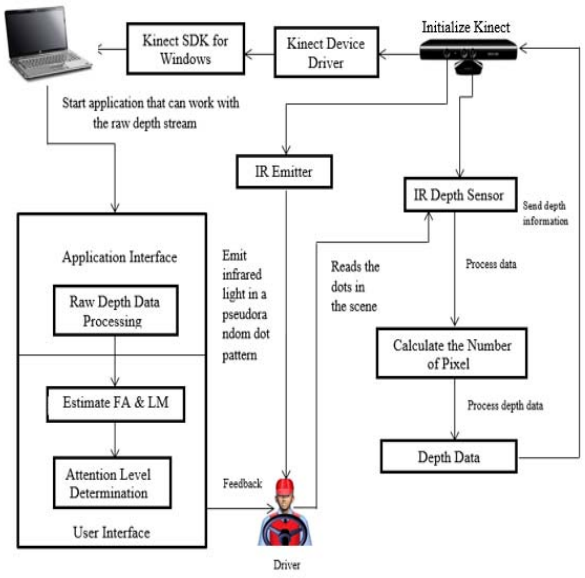

Fig 1: Schematic framework to estimate the driver's level of attention

B. Identification of driver and tracking

We used a model based method for human detection from depth images. 2-stage head detection process was used for human detection, which includes a 2D edge detector and a 3D shape detector. For a given input human a depth array is generated, reduce noise and smooth the array for later process. We first explore the boundary information embedded in the depth array to locate the candidate regions that may indicate the appearance of people. The algorithm used here is 2D chamfer distance matching. It scans across the whole image and gives the possible regions that may contain people. We examine each of these regions using a 3D head model, which utilizes the relational depth information of the array for verification. We extract the parameters of the head from the depth array and use the parameter to build a 3D head model. Then we match the 3D model against all the detected regions to make a final estimation. From the detection result of 2D chamfer matching, we can get the depth of the head like object from the depth array.

In order to generate virtual 3D model of sphere (i.e. head), the only parameter that is required is radius of the sphere has conducted a regression test that have arrived at a cubic equation that gives the relationship between the depth value and the approximate height of head. The cubic equation (1) that fits the geometrical height of the head is computed as

$$-y = -1.3835\times10^{-9}x^3+1.8435\times10^{-5}x^2-0.091403x+189.38 \quad (1)$$

Calculate the standard height of the head in this depth by (1). Then search for the head within a certain range that is defined by the standard height of the head by (2).

(2)

Here, h is the height of the head calculated from equation (1), R is the search radius. Now, we need to convert the height from millimeter into pixel unit as the matching works pixel by pixel.

Calculate the radius of the head as (3).

$$\text{Radius (pixel)} = \left(\frac{1}{1.2}\right) \times R \qquad (3)$$

Fit the model onto the regions detected from previous steps. We extract a circular region with radius R around the detect center and normalize its depth in (4).

$$Err = \sum | \text{normalized\_real\_depth}(i, j) - \text{ideal\_depth}(i, j) |^2 \qquad (4)$$

We used a threshold value to decide the region is actually a head or not. We also remove the false positives. If the matching value is within the threshold, then we say the head is detected. Detection of head means there is a human. When the driver's head is detected, we track the face using Microsoft Software Development Kit for Kinect for Windows (Face Tracking SDK), together with the Kinect for Windows Software Development Kit (Kinect for Windows SDK) in real time [14].

We give preliminary results on tracking using depth information based on our detection result. Our tracking algorithm is based on the movements of human head. We assume that the coordinates and speed of the same objects in neighboring frames change smoothly or randomly. Then by using, the infrared (IR) depth camera persons in the field of view of the camera can be tracked. The head points of the tracked users can be located in space and their head movements be tracked. The work flow of the human head tracking system [11] is presented in a flowchart as in Fig. 2. Users can be detected either while standing or sitting and facing the Kinect. Yaw angle is used to estimate the driver's attention level.

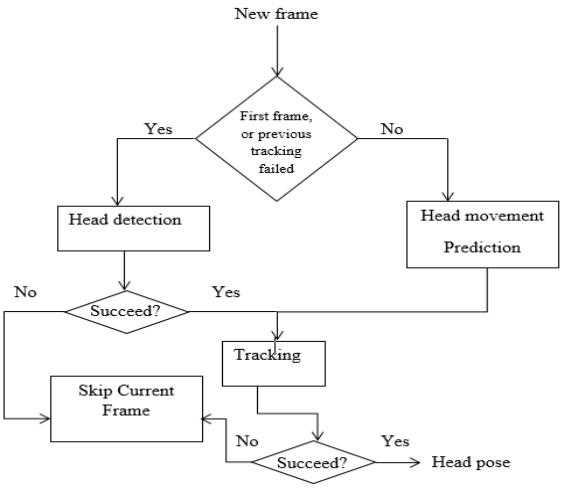

Fig. 2: Workflow of the Human Head Tracking System

C. Depth information processing

Then IR emitter constantly emits infrared light in a random dot pattern over everything in front of it. These dots are normally invisible to us, but it is possible to capture their depth information using an IR depth sensor. The dotted light reflects off different objects. IR depth sensor that reads the dots in the scene, processes the data, and sends the depth information from which they were reflected. IR depth sensor starts reading the inferred data from the object based on the distance of the individual light points of reflection IR depth sensor then passes it to the prime sense chip within the device. The prime sense chip then analyzes the captured data and produces a depth image. The Kinect sensor then returns the depth stream data as a succession of the depth image frame. The Kinect sensor returns raw depth with 16-bit grey scale format.

D. Measuring the level of attention

Attention can be considered as a function of two parameters face angle (FA) and lip motion (LM),

$$L(A) = f(FA, LM)$$

Where, L(A) is the level of attention of the driver, FA is the facial angle respect to Kinect and LM represents the lip motion of the driver respectively.
Attention depends on facial angle (FA) inverse proportionally. Lower angle indicates the higher level of attention and vice versa. Similarly, attention also depending on the lip motion (LM).

$$\mathbf{A_{fa}} = \begin{cases} 1/\{FA\} & ; \text{when } FA>0 \\ 1/\{FA * (-1)\} & ; \text{when } FA<0 \end{cases} \qquad (5)$$

$$\mathbf{A_{lm}} = 1/\{LM\} \qquad (6)$$

$$\mathbf{A} = (A_{fa} + A_{lm}) \qquad (7)$$

Here, $\mathbf{A_{fa}}$ is the attention with respect to aace angle (FA) and $\mathbf{A_{lm}}$ is the attention with respect to lip motion (LM) of the driver. Level of attention is determined for the obtained $\mathbf{A_{fa}}$ and $\mathbf{A_{lm}}$ using (7).

E. Range of level pf attention

LA is defined according to the number of predefined ranges of attention level. We convert the values of interest into percentage. Rule for scaling into percentage as in (8),

$$(A-A_{min})/(A_{max}-A_{min}) = (A_p-A_{pmin})/(A_{pmax}-A_{pmin}) \qquad (8)$$

Where A = Level of attention found from equation (7), A_{min} = minimum LI generated by the system, A_{max} = maximum LI generated by the system, Ap = I in percentage, A_{pmin} = 0 (minimum value in percentage), A_{pmax} = 100 (maximum value in percentage). The range of level of attention is listed in

Table 1. If the driver's attention gradually goes down to the range of Very Low Attention, then the system will alarm the driver to track back to the range of Maximum level of attention.

TABLE I: RANGE OF LA

Range of Attention (%)	**Level of Attention (LA)**
If A == 0 then,	No Attention
If A < 10 then,	Minimum Attention
If A < 20 then,	Very Low Attention
If A < 30 then,	Low Attention
If A < 45 then,	Lower Medium Attention
If A < 55 then,	Medium Attention
If A < 70 then,	Higher Medium Attention
If A < 80 then,	High Attention
If A < 90 then,	Very High Attention
If A < = 100 then,	Maximum Attention

IV. EXPERIMENTAL RESULTS

We recruited 10 drives in the real vehicle environment to run the experiment. The average age of participants are 25 years (SD= 4.50). In this experiment to get the appropriate attention level of driver, we positioned different people in the driving seat within the device range of area. Participants were interacted with the system one by one.

A. Experimental setup

Experimental Environment is created by connecting Kinect with the computer via power adapter and also set the sensor device in a proper position in front of the driver. During testing, the driver has to be seated directly in front of the system environment which is established with Microsoft Kinect and computer. The laptops built in keyboard and mouse are used for manipulate the Application Interface. The Kinect sensor was mounted at a distance of 3 inches above the steering of the vehicle and facing directly at the subject (Driver). For this experiment the Kinect Studio was used to record the RGB and depth stream for later processing.

B. Graphical User Interface

GUI is developed to observe the measured attention level of the driver as well as values of two parameters face angle and lip motion. We can see the total graphical User Interface of our system with participant in Fig 3. It also represent graphical representation of these measured values in real time, which is very important to realize the variation of the attention level over time and also important to alert the driver if the attention level critically falls down through alarm. Immediately after a human is found, the human is detected by the system sensors and his face is tracked by a mask of mesh. After detection, the system records depth data of that part in real-time and tracks the driver's face.

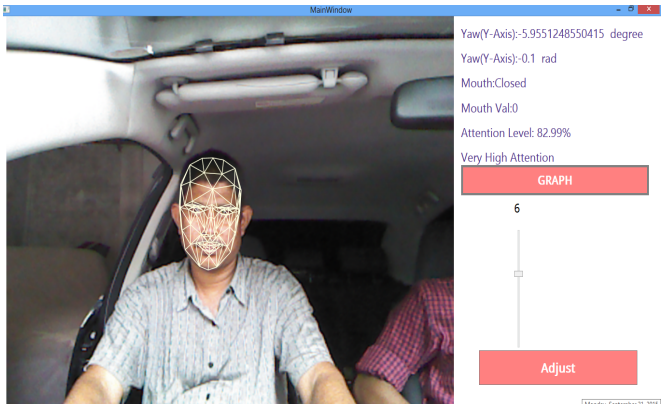

Fig. 3: Total graphical interface of our system with subject

Participants were asked to ride the car slowly as well as to move their face to left and right randomly and speak to detect the lip motion to evaluate the functionality of detection and tracking process. It also tests whether the system is resulting real-time change of face angle, lip motion and level of attention. Each trial was started with a fixed positioning of a participant. Each of the participants attended more than 10 trials for the evaluation of each of the module separately. It took approximately 2-3 hours to perform the experiment.

C. Visualization

It is clearly seen that the level of attention is changing inversely proportional to the value of combined parameter (i.e. face angle and lip motion). This change in the graphical view is considered as one of the important evaluation factors. In all the above cases we see that the graphical view extremely matches the mathematical and conceptual views. So it can be said that the experiment results in a quite satisfactory performance. The complete interface of the graphical representation of the attention level of the driver is illustrated in Fig. 4.

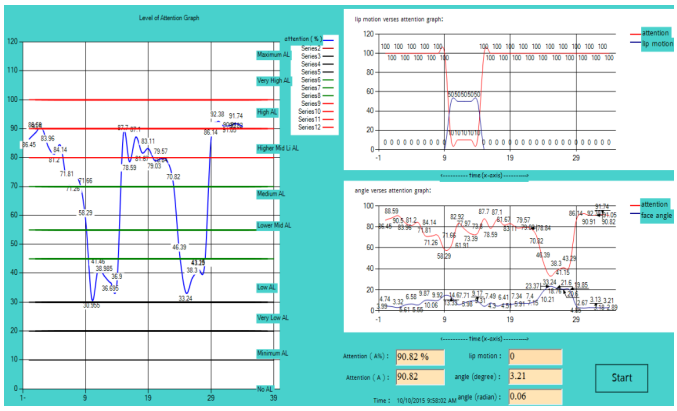

Fig. 4: Complete interface of graphical representation of LA

D. Evaluation of FA estimation module

We positioned the participant in driving seat within the viable range and more than 10 trials have been given by each of the participants. The qualitative evaluation of the module, which is performed with the experimental data are given in Table 2.

TABLE 2: PERFORMANCE EVALUATION OF FA MODULE

	Actual LM	LM Detected	Failed	Accuracy (%)	Error Rate (%)
Participant 1	12	10	2	83.33%	16.67%
Participant 2	15	14	1	93.33%	6.67%
Participant 3	10	10	0	100%	0%
Participant 4	14	12	2	85.71%	14.28%
Participant 5	8	7	1	87.5%	12.5%
Participant 6	16	15	1	93.75%	6.25%
Participant 7	10	9	1	90%	10%
Participant 8	15	13	2	86.66%	13.34%
Participant 9	10	10	0	100%	0%
Participant 10	13	12	1	92.30%	7.7%

E. Evaluation of LM estimation module

We evaluate the performance of lip motion (LM) on the basis of the comparison of physical and system detecting lip motion of the driver. The participants are told to move their lip randomly while sitting in front of the Kinect. Fig. 5 shows the graphical representation of performance evaluation of LM module.

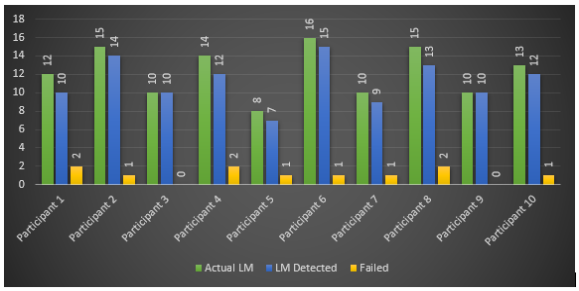

Fig. 5: Performance analysis of lip motion estimation module

Table III shows the results of analysis of LM estimation module.

TABLE III: PERFORMANCE EVALUATION OF LM MODULE

	No of Trials	Estimate FA	Deviation	Accuracy (%)	Error Rate (%)
Participant 1	10	9	1	90%	10%
Participant 2	12	10	2	83.33%	16.67%
Participant 3	15	13	2	86.66%	13.37%
Participant 4	8	8	0	100%	0%
Participant 5	13	11	2	84.61%	15.39%
Participant 6	10	9	1	90%	10%
Participant 7	12	12	0	100%	0%
Participant 8	15	14	1	93.33%	6.67%
Participant 9	13	10	3	76.92%	23.08%
Participant 10	14	13	1	92.85%	7.15%

F. Attention level summarization

First calculate the level of attention (LA) against face angle and lip motion individually. Then combined both parameters and calculated the level of attention. The summarization of level of attention is presented below in Table IV.

TABLE IV: SUMMARIZATION OF LEVEL OF ATTENTION ESTIMATION

Case No.	Face Angle (FA)	Lip Motion (LM)	LA for Single parameter (FA)	IL for Single parameter	LA for double Parameter (FA + LM)	Attention Level
1	0.34	No	44.85%	100%	44.85%	LMA
2	0.10	No	83%	100%	82.99%	VHA
3	0.08	Yes	87.4%	0%	43.70%	LMA
4	0.17	Yes	71.61%	0%	35.73%	LMA
5	0.20	Yes	67.28%	0%	33.64%	HMA
6	0.41	No	23.5%	100%	23.5%	LMA
7	0.36	No	41.68%	100%	41.66%	LMA
8	0.43	Yes	30%	0%	14.99%	VLA
9	0.08	No	86.57%	100%	86.57%	VHA
10	0.21	Yes	65.37%	0%	32.68%	LWA
Average			60.12%	50%	44.03%	LMA

Fig. 6 shows the graphical representation of attention with respect to both parameters (FA+LM).

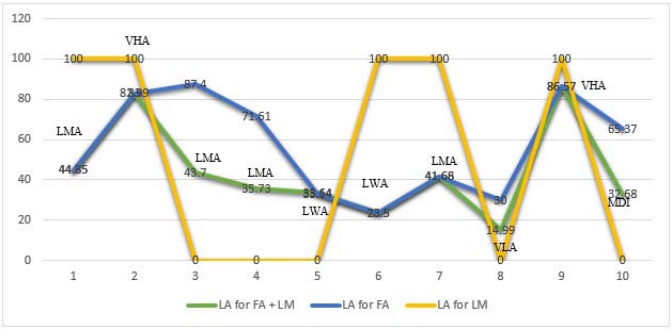

Fig. 6: Graphical representation of attention based on FA, LM and (FA+LM) respectively.

Figure 6 indicates that, the attention graph with respect both parameters is changes moderately than attention based on individual parameter. When the attention level with respect to LM is 100% then the line representing the attention level for FA is super positioned on the line representing the attention level for (FA+LM). Therefore, we can conclude that combined parameters produced the better results for level of attention estimation.

G. Alarm generating module

There may a possibility of vehicle crashes or severe road accident when the level of attention is low. For this reason it is very much important to aware the driver when the attention level is very low for a while. Proposed system generates an alarming sound to alert the driver when his/her attention is below 20% i.e. very low. If the driver is not being attentive after listening the alarm, then the system will generate continuous sound to draw the driver's attention until the driver's high attention level is assured. Fig. 7 shows a snapshot for alarm generating module.

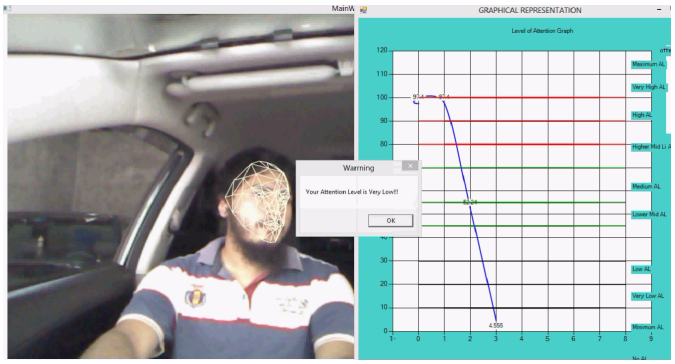

Fig. 7: Generating alarm with a message box to alert the driver

In Fig. 7, we can see a message box is generated with an alarming sound when the attention level of the driver goes down to 20%.

V. CONCLUTION

This paper proposes a vision based framework for measuring the level of attention of the driver based on face angle and lip motion parameters. As we worked with depth data rather than that of RGB image processing, performance is quite improved overcoming the environmental effects. Image based processing requires complex algorithm and complex computational costs. We used simpler algorithm for dealing with depth data. This definitely contributes to speed up the overall system performance. If the system can measure the level of attention the driver properly and generate alarm when the driver's attention level becomes very low then it will be a shield to reduce the high accident rate as well as to save the human life from the curse of road accident. This is a very significant fact in the field of safe driving. More physiological parameters such as alcohol detector, drowsiness detector may be considered in future to improve the performance of the system.

References

[1] F. J. P. Alcalde, I. G. Varea and J. M. G´omez, "Active attention for human robot interaction using visual and depth information," available: http://neithan.weebly.com, 2012.

[2] Y. Sun and R. Fisher, "Object-based visual attention for computer vision, Artificial Intelligence," vol. 146, no. 1, pp. 77–123, 2003.

[3] D. Das, Y. Kobayashi and Y. Kuno "Attracting attention and establishing a communication channel based on the level of visual focus of attention," in 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2194-2201, 2013.

[4] L. Li, X. Yu, J. Li, G. Wang, J. Y. Shi, Y. K. Tan, and H. Li "Vision-based attention estimation and selection for social robot to perform natural interaction in the open world," in Institute for Infocomm Research, pp.183-184, 2012.

[5] S. O. Ba, J. M. Odobez "Recognizing visual focus of attention from head pose in natural meeting," vol. 39, pp. 16-33, 2008.

[6] K. Lai, J. Konrad, P. Ishwar "A gesture-driven computer interface using Kinect," in Image Analysis and Interpretation (SSIAI), pp. 185-188, 2012.

[7] R. A. El-laithy, J. Huang, M. Yeh "Study on the use of Microsoft Kinect for robotic applications," in Position Location and Navigation Symposium (PLANS), pp. 1280-1288, 2012.

[8] M. Kinect, "Kinect face tracking," https://msdn.microsoft.com/en-us/library/jj130970.aspx, 2013.

[9] A. Jana, "Kinect for windows SDK Guide," [online]. Available: www.it-ebooks.infos, 2012.

[10] S. Li, K. N. Ngan and L. Sheng "A head pose tracking system using RGB-D camera," in ICVS, pp. 153-162, 2013.

[11] A. Yargic, M. Dogan, "A Lip Reading Application on Kinect Camera," in Innovations in Intelligent System and Application (INISTA), pp. 1-5, 2013.